C L A I M S

What is claimed is:

1. A method of clustering a set of records, each of the records having attribute values for a set of attributes, the method comprising:

for each attribute of the set of attributes, determining a characteristic value for said each attribute, based on attribute values of said each attribute;

for each attribute value, determining a deviation from the characteristic value of said each attribute;

for each record, sorting the set of attributes based on deviations of the attribute values, to provide a key; and

clustering the set of records based on the key.

2. The method of claim 1, further comprising calculating a mean value of the attribute values of said each attribute as the characteristic value.

3. The method of claim 1, wherein a median value of the attribute values of said each attribute is determined as the characteristic value.

4. The method of claim 1, wherein determining the deviation comprises calculating a difference between said each attribute value and the characteristic value of said each attribute.

5. The method of claim 1, wherein determining the deviation comprises calculating a difference between said each attribute value and the characteristic value of the corresponding attribute, and dividing the difference by the characteristic value of said each attribute.

6. The method of claim 1, wherein sorting the set of attributes comprises using absolute values of the deviations of the attribute values as a sorting criterion.

7. The method of claim 1, wherein a first record of the set of records contains a first key and a second record of the set of records contains a second key; and

further comprising placing the first key and the second key into a single cluster if the first key and the second key have identical sub-sequences of a first length.

8. The method of claim 1, wherein a first record of the set of records contains a first key and a second record of the set of records contains a second key; and

further comprising placing the first key and the second key into a single cluster if the first key and the second key have identical sub-sequences of absolute values of the deviations.

9. The method of claim 1, wherein a first record of the set of records contains a first key that has a first sub-sequence, and a second record has a second sub-sequence contains a second key; and

further comprising placing the first key and the second key into a single cluster if the first and second sub-sequences comprise the same set of attributes.

10. The method of claim 9, wherein the first and second sub-sequences comprise the same set of attributes irrespective of a sign of the deviations of the attribute values.

11. The method of claim 10, further comprising:

identifying a cluster having a smallest number of records; and

for each record of the identified cluster searching another cluster having records with best matching keys.

12. The method of claim 11, further comprising reducing a length of the first sub-sequence and a length of the second sub-sequence in order to find a best match.

13. The method of claim 12, further comprising using a distance measure to find another cluster for a record of the identified cluster.

14. The method of claim 13, wherein the distance measure comprises a Euclids distance.

15. A computer program product having instruction codes for clustering a set of records, each of the records having attribute values for a set of attributes, the computer program product comprising:

a first set of instruction codes, which, for each attribute of the set of attributes, determines a characteristic value for said each attribute, based on attribute values of said each attribute;

a second set of instruction codes, which, for each attribute value, determines a deviation from the characteristic value of said each attribute;

a third set on instruction codes, which, for each record, sorts the set of attributes based on deviations of the attribute values, to provide a key; and

a fourth set of instruction codes for clustering the set of records based on the key.

16. The computer program product of claim 15, further comprising a fifth set of instruction codes for calculating a mean value of the attribute values of said each attribute as the characteristic value.

17. The computer program product of claim 15, further comprising a sixth set of instruction codes for setting a median value of the attribute values of said each attribute as the characteristic value.

18. The computer program product of claim 15, wherein the second set of instruction codes determines the deviation by calculating a difference between said each attribute value and the characteristic value of said each attribute.

19. The computer program product of claim 15, wherein the second set of instruction codes determines the deviation by calculating a difference between said each attribute value and the characteristic value of the corresponding attribute, and by dividing the difference by the characteristic value of said each attribute.

20. The computer program product of claim 15, wherein the third set on instruction codes sorts the set of attributes using absolute values of the deviations of the attribute values as a sorting criterion.

21. A system for clustering a set of records, each of the records having attribute values for a set of attributes, the system comprising:

each attribute of the set of attributes comprising a characteristic value for said each attribute based on attribute values of said each attribute;

each attribute value comprising a deviation from the characteristic value of said each attribute;

each record comprising the set of attributes based on deviations of the attribute values, to provide a key; and

wherein the set of records are clustered based on the key.

22. The system of claim 21, wherein a mean value of the attribute values of said each attribute is calculated as the characteristic value.

23. The system of claim 21, wherein a median value of the attribute values of said each attribute is calculated as the characteristic value.

24. The system of claim 21, wherein the deviation is calculated as a difference between said each attribute value and the characteristic value of said each attribute.

25. The system of claim 21, wherein the deviation is determined by calculating a difference between said each attribute value and the characteristic value of the corresponding attribute, and by dividing the difference by the characteristic value of said each attribute.

26. The system of claim 21, wherein the set of attributes is sorted using absolute values of the deviations of the attribute values as a sorting criterion.